American Journal of Epidemiology
Vol. 166, No. 8
DOI: 10.1093/aje/kwm248
Advance Access publication September 5, 2007

Editorial

Turning the Pump Handle: Evolving Methods for Integrating the Evidence on Gene-Disease Association

Julian P. T. Higgins[1], Julian Little[2], John P. A. Ioannidis[3,5], Molly S. Bray[4], Teri A. Manolio[6], Liam Smeeth[7], Jonathan A. Sterne[8], Betsy Anagnostelis[9], Adam S. Butterworth[10], John Danesh[10], Carol Dezateux[11], John E. Gallacher[12], Marta Gwinn[13], Sarah J. Lewis[8], Cosetta Minelli[14], Paul D. Pharoah[15], Georgia Salanti[3], Simon Sanderson[10], Lesley A. Smith[16], Emanuela Taioli[17], John R. Thompson[18], Simon G. Thompson[1], Neil Walker[19], Ron L. Zimmern[20], and Muin J. Khoury[13]

[1] MRC Biostatistics Unit, Cambridge, United Kingdom.
[2] Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa, Ontario, Canada.
[3] Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece.
[4] Center for Human Genetics, Institute of Molecular Medicine and School of Public Health, University of Texas, Houston, TX.
[5] Department of Medicine, Tufts University School of Medicine, Boston, MA.
[6] National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.
[7] Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom.
[8] Department of Social Medicine, University of Bristol, Bristol, United Kingdom.
[9] Royal Free Hospital Medical Library, University College London, London, United Kingdom.
[10] Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom.
[11] Centre for Paediatric Epidemiology and Biostatistics, Institute of Child Health, University College London, London, United Kingdom.
[12] Department of Epidemiology, Cardiff University, Cardiff, Wales, United Kingdom.
[13] National Office of Public Health Genomics, Centers for Disease Control and Prevention, Atlanta, GA.
[14] National Heart and Lung Institute, Imperial College, London, United Kingdom.
[15] Cancer Research UK Human Cancer Genetics Group, Department of Oncology, University of Cambridge, Cambridge, United Kingdom.
[16] School of Health and Social Care, Oxford Brookes University, Oxford, United Kingdom.
[17] University of Pittsburgh Medical Center, Pittsburgh, PA.
[18] Department of Health Sciences, University of Leicester, Leicester, United Kingdom.
[19] Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, United Kingdom.
[20] PHG Foundation, Cambridge, United Kingdom.

Received for publication July 2, 2007; accepted for publication August 1, 2007.

Abbreviations: HuGE, Human Genome Epidemiology; HuGENet, Human Genome Epidemiology Network.

Editor's note: *This paper is also available on the website of the Human Genome Epidemiology Network (http://www.cdc.gov/genomics/hugenet/).*

Recent findings from genome-wide association studies have demonstrated their considerable potential for identifying genetic determinants of common diseases of public health significance such as cancer, heart disease, and diabetes (1), but they have also highlighted the continued importance of targeted genotyping to replicate genome-wide association findings (2). Approaches to the integration of evidence in human genome epidemiology have evolved rapidly in the last few years. The combination of results from

Correspondence to Dr J. P. T. Higgins, MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 0SR, United Kingdom (e-mail: julian.higgins@mrc-bsu.cam.ac.uk).

multiple studies, often known as meta-analysis, has a key role both in enhancing power and in characterizing relative risks (3). As evidence accumulates on genetic variants that confer identifiable effects on disease susceptibility, so does the need to summarize the evidence in digestible and accessible formats. Here, we describe how the Human Genome Epidemiology Network (HuGENet) is keeping abreast of developments in methods for collating and synthesizing the evidence.

HuGENet was established in 1998 to integrate epidemiologic evidence on the role of genetics in human health and disease, and to develop an online searchable, updated, knowledge base (4). HuGENet's main activities are compilation and evaluation of epidemiologic research, facilitating of collaborations, training and technical assistance, and information exchange through the World Wide Web. A ''road map'' for human genome epidemiology outlines a vision for the future of this important field (5), and activities of the network are now facilitated by four coordinating centers in Atlanta, Georgia (6); Cambridge, United Kingdom (7); Ottawa, Canada (8); and Ioannina, Greece (9).

An important part of the HuGENet initiative is conducting ''Human Genome Epidemiology (HuGE) reviews'' on genotype-disease associations, including joint effects of genes and of genes with environmental exposures (10, 11). Indeed, HuGENet's new logo (figure 1) highlights the central role of gene-environment (GE) interactions in predisposition to disease. HuGE reviews are typically *systematic*, aiming to identify, appraise, and synthesize evidence from all relevant existing studies on the topic in question (12). Regular readers will have noticed an increasing number of HuGE reviews in the *American Journal of Epidemiology*, and nine additional journals have agreed to be publication venues for these systematic reviews (6). HuGE reviews may also be accessed from the HuGENet website; the 62 HuGE reviews published as of June 1, 2007, have covered a wide array of topics ranging from rare, single-gene disorders such as neurofibromatosis to common conditions such as preterm birth, cancer, and heart disease (6).

The *HuGE Review Handbook* (13) is an evolving, online document that offers guidance to researchers undertaking HuGE reviews. It is inspired partly by the *Cochrane Handbook for Systematic Reviews of Interventions* (14). The Cochrane Collaboration undertakes systematic reviews of the effects of health-care interventions and has published more than 2,500 such reviews to date (15). Cochrane reviews implement rigorous methods in an attempt to minimize bias either from individual studies or during the review process, and similar rigor is being used in HuGE reviews. The *Handbook* will be updated over time as methodology and understanding develop.

The *Handbook* resulted from a methodology workshop held in Cambridge, United Kingdom, in November 2004. The workshop brought together epidemiologists, geneticists, statisticians, and other health-care researchers to develop methodological guidance for authors of systematic reviews and meta-analyses in human genome epidemiology, and to identify any potential developments that could improve their validity. Before this workshop, the original (4) and updated (11) guidelines for HuGE reviews did not specify in detail the recommended methods for searching published and unpublished literature, analyzing data, or synthesizing information. Furthermore, the initial concept of a ''full'' HuGE review—to cover prevalence, association, interaction, and implications for genetic testing and public health (16)—was relatively broad in scope. Thus, early HuGE reviews varied in their methodology and particularly in the application of formal meta-analytic methods. This reflected concern about the application of meta-analysis to observational studies (16, 17). However, meta-analysis has become widely applied and accepted in human genome epidemiology in recent years (18, 19), and all but four of the 17 HuGE reviews published since the beginning of 2006 include a formal meta-analysis. Furthermore, over 500 meta-analysis articles have already been published in this field outside the HuGENet effort (6). With the advent of genome-wide association studies, it has become common practice that prospective validation of identified variants through combined analysis (or meta-analysis) of data from multiple teams is accomplished as part of the very first publication of the new data (3). Meta-analysis of genome-wide association studies themselves is also increasingly applied (20, 21).

FIGURE 1. The HuGENet logo.

Some key recommendations in the *HuGE Review Handbook* for improving the methodology of HuGE reviews include the following:

1. Encouraging consortia of primary research investigators as the most reliable approach for performing combined analyses or meta-analyses (based on individual participant data) (22)
2. Adopting methods to minimize human error in the literature-based reviews, such as duplicating selection of studies and data extraction
3. Conducting comprehensive (yet practically realistic) searches for eligible studies, considering sources beyond MEDLINE (National Library of Medicine, Bethesda, Maryland)
4. Considering in more detail the potential for bias in individual studies and in the total body of available evidence (17)
5. Encouraging quantitative synthesis of results from multiple studies (meta-analysis) where appropriate (23)

6. Encouraging incorporation of intermediate phenotypes (such as molecular markers) so that ''Mendelian randomization'' can be exploited to examine the causal effects of such phenotypes (24)

Meta-analyses can offer both enhanced power to detect associations and increased precision of estimates of its magnitude. Consistency of findings across studies can be formally assessed and heterogeneity explored. Of course, the potential for selective availability of findings on the basis of their statistical significance must always be borne in mind. It is essential that the scientific community continues to progress toward making all findings, positive and negative, available to all. Registers of DNA collections (akin to existing registers of randomized controlled trials (25)) and online repositories for negative results would go some way toward realizing the vision of an unbiased and data-rich environment within which to evaluate gene-disease associations. Genome-wide association investigations offer a unique opportunity for full, transparent availability of detailed databases to other researchers, such as those already adopted by the Wellcome Trust Case Control Consortium (26), the National Institutes of Health's Database of Genotype and Phenotype (dbGaP) (27), and the European Genotype Archive (28). Wherever possible, we encourage the development of consortia of investigators to analyze individual participant data on at least a retrospective basis and, ideally, a prospective basis.

The fast pace of development in the field creates new challenges such as the need to continually revisit the inferences of meta-analyses and the definition of replication in the context of massive testing ability (29, 30). Inferences on the cumulative evidence on genetic associations may change over time. As part of a HuGENet initiative, interim guidelines have been developed to assess the epidemiologic strength of the cumulative evidence (31). We suggest that these guidelines be applied to the final inference for each HuGE review and other meta-analyses or field synopses (6). We expect the *Handbook* to be a dynamic enterprise that will be regularly updated to recognize consensus on best methods for new challenges, as these arise. We encourage others interested in human genome epidemiology to contribute to this process and to contact us via one of the websites listed below:

HuGENet United States (main site): www.cdc.gov/genomics/hugenet

HuGENet Canada: www.hugenet.ca

HuGENet Greece: www.dhe.med.uoi.gr/hugenet.htm

HuGENet United Kingdom: www.hugenet.org.uk

A decade ago, Shpilberg et al. declared that ''the sequencing of the human genome offers the greatest opportunity for epidemiology since John Snow discovered the Broad Street Pump'' (32, p. 637). With the completion of the Human Genome Project in 2003, the era of developing the handle for the pump formally began (16). With the emergence of new tools now available to epidemiologists, we hope the pump handle will begin to turn, albeit slowly, to uncover the secrets of gene-environment interactions in common human diseases.

ACKNOWLEDGMENTS

The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

Conflict of interest: none declared.

REFERENCES

1. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661–78.
2. Todd JA, Walker NM, Cooper JD, et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat Genet 2007;39:857–64.
3. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007;447:1087–93.
4. Khoury MJ, Dorman JS. The Human Genome Epidemiology Network. Am J Epidemiol 1998;148:1–3.
5. Ioannidis JPA, Gwinn M, Little J, et al. A road map for efficient and reliable human genome epidemiology. Nat Genet 2006;38:3–5.
6. CDC National Office of Public Health Genomics. Human Genome Epidemiology Network. (http://www.cdc.gov/genomics/hugenet). (Accessed June 1, 2007).
7. MRC Biostatistics Unit, Cambridge. HuGENet UK Coordinating Centre. (http://www.hugenet.org.uk). (Accessed June 1, 2007).
8. Department of Epidemiology and Community Medicine, University of Ottawa. HuGENet Canada Coordinating Centre. (http://www.hugenet.ca). (Accessed June 1, 2007).
9. Department of Hygiene and Epidemiology, University of Ioannina School of Medicine. HuGENet. (http://www.dhe.med.uoi.gr/hugenet.htm). (Accessed June 1, 2007).
10. Khoury MJ, Little J. Human genome epidemiologic reviews: the beginning of something HuGE. Am J Epidemiol 2000;151:2–3.
11. Revised guidelines for submitting HuGE reviews. Am J Epidemiol 2000;151:4–6.
12. Egger M, Davey Smith G, Altman DG, et al. Systematic reviews in health care: meta-analysis in context. London, United Kingdom: BMJ, 2001.
13. Little J, Higgins J, eds. HuGENet HuGE review handbook. (http://www.hugenet.ca). (Accessed June 1, 2007).
14. Higgins JPT, Green S, eds. Cochrane handbook for systematic reviews of interventions 4.2.6 (updated September 2006). In: The Cochrane Library, issue 4. Chichester, United Kingdom: John Wiley & Sons, Ltd, 2006.
15. The Cochrane Collaboration. Cochrane Database of Systematic Reviews. In: The Cochrane Library. Chichester, United Kingdom: John Wiley & Sons, Ltd, 2006.
16. Little J, Khoury MJ, Bradley L, et al. The human genome project is complete. How do we develop a handle for the pump? Am J Epidemiol 2003;157:667–73.
17. Little J, Bradley L, Bray MS, et al. Reporting, appraising, and integrating data on genotype prevalence and gene-disease associations. Am J Epidemiol 2002;156:300–10.

18. Ioannidis JPA, Ntzani EE, Trikalinos TA, et al. Replication validity of genetic association studies. Nat Genet 2001;29:306–9.
19. Lohmueller KE, Pearce CL, Pike M, et al. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 2003;33:177–82.
20. Zeggini E, Weedon MN, Lindgren CM, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 2007;316:1336–41.
21. Evangelou E, Maraganore DM, Ioannidis JPA. Meta-analysis in genome-wide association datasets: strategies and application in Parkinson disease. PLoS ONE 2007;2:e196.
22. Ioannidis JPA, Bernstein J, Boffetta P, et al. A network of investigator networks in human genome epidemiology. Am J Epidemiol 2005;162:302–4.
23. Salanti G, Sanderson S, Higgins JPT. Obstacles and opportunities in meta-analysis of genetic association studies. Genet Med 2005;7:13–20.
24. Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? Int J Epidemiol 2003;32:1–22.
25. De Angelis C, Drazen JM, Frizelle FA, et al. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. Lancet 2004;364:911–12.
26. The Wellcome Trust Case Control Consortium. (http://www.wtccc.org.uk). (Accessed June 1, 2007).
27. National Library of Medicine (National Institutes of Health). Database of Genotype and Phenotype (dbGaP). (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gap). (Accessed June 1, 2007).
28. The European Genotype Archive. (http://www.ebi.ac.uk/ega). (Accessed June 1, 2007).
29. Chanock SJ, Manolio T, Boehnke M, et al. Replicating genotype-phenotype associations. Nature 2007;447:655–60.
30. Ioannidis JPA. Non-replication and inconsistency in the genome-wide association setting. Hum Hered 2007;64:203–13.
31. Ioannidis JP, Boffetta P, Little J, et al. Assessment of cumulative evidence on genetic associations: interim guidelines. Int J Epidemiol (in press).
32. Shpilberg O, Dorman JS, Ferrell RE, et al. The next stage: molecular epidemiology. J Clin Epidemiol 1997;50:633–8.